

RELLENO DE DATOS DE PRECIPITACIÓN Y TEMPERATURA MEDIA DE LA ESTACIÓN METEOROLÓGICA M0151-ZAPOTILLO

Fundamento. El Instituto Nacional de Meteorología e hidrología (INAMHI) es el ente rector de la meteorología e hidrología en el país, por tal motivo y, considerando que existe un sinnúmero de proyectos relacionados al medio ambiente que de una u otra manera pueden ocasionar impactos de distinto índole, necesitan de datos climáticos que describan el comportamiento temporal y espacial de las distintas variables meteorológicas; por tal motivo es primordial considerar la información meteorológica que genera esta Entidad del Estado, lamentablemente con el pasar de los años la misma ha ido decreciendo tanto en cantidad como en cantidad por lo que las series históricas han sufrido serias rupturas de información; por tal motivo se ha procedido a realizar el relleno de las series de precipitación y temperatura media en base a los datos disponibles en función del método estadístico determinado.

Objetivo.

General: Rellenar las series temporales de precipitación y temperatura media del período 1991-2020 mediante el método estadístico conocido como Método de Regresión Ortogonal ya que es un método mediante el cual se pueden obtener valores más precisos a la realidad.

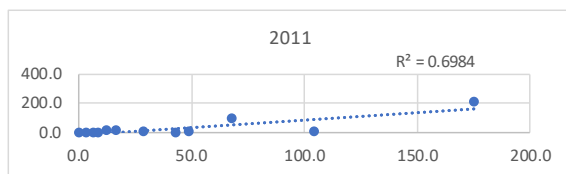
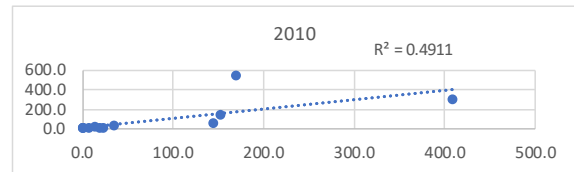
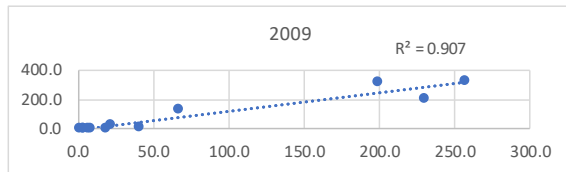
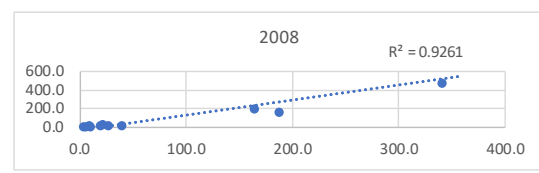
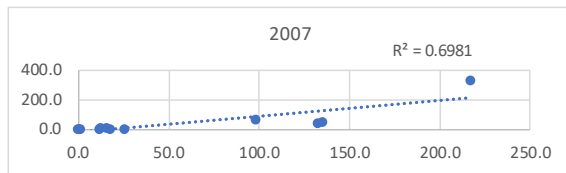
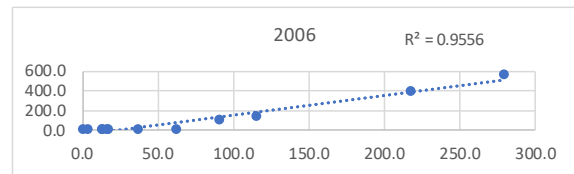
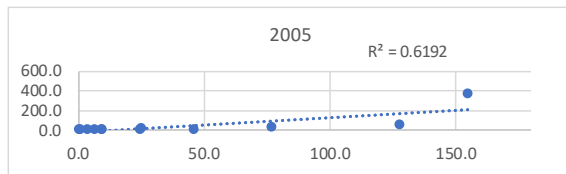
Metodología: La información proporcionada por el INAMHI de la estación meteorológica M0151 – Zapotillo tiene datos a partir de 1980 misma que hasta el año 2011 es relativamente continua, a partir del año 2012 es muy notoria la falta de información en el transcurso de un mismo año calendario sino que también durante casi cinco años consecutivos como lo es entre el período 2016 al 2020, en tal sentido y, en vista de que en la zona no existen otras estaciones meteorológicas, se ha procedido a obtener datos generados de una Interfaz de Programación de Aplicaciones (API), misma que utiliza información producto del reanálisis de los más recientes modelos meteorológicos y los actualmente más avanzados y desarrollados. Esta API esta bajo el dominio de Open-Meteo¹.

Con el propósito de obtener normales climatológicas tanto de precipitación como de temperatura media que permitan entender su variación interanual, en un principio se procedió a contabilizar el porcentaje de información disponible para el período 2001-2020 dando como resultado un 62.1% de datos disponibles y consecuentemente el 37.9% de datos faltantes, esto en lo que respecta a la variable de precipitación, mientras que para la temperatura media el porcentaje con datos disponibles es del 50.8% y del 49.2% de ausencia de los mismos; en el período 1991-2020 el porcentaje de datos registrados se incrementa para la variable de precipitación ubicándose en el 71.1%, el restante 28.9% corresponde a la carencia de los mismos; en cuanto a la temperatura media estos son del 56.9% y 43.1% respectivamente. En base a esta consideración se optó por utilizar la serie 1991-2020 ya que la cantidad de datos faltantes para ser rellenados es menor tanto para la precipitación como para la temperatura media.

Desde el punto de vista técnico no es aconsejable realizar relleno de datos cuando en la serie en cuestión se supera el 25% de datos faltantes pero debido a la necesidad de contar con esta información se procede a realizar pruebas para determinar si los datos obtenidos a través de la API con los registrados en la estación Zapotillo se correlacionan entre sí.

Es así que se procedió a realizar gráficos de dispersión para determinar si existe o no relaciones complementadas con líneas de tendencia lineal entre las dos series de datos consideradas para los años del 2005 al 2011 por ser aquellos que tienen registros completos durante todo el año calendario, el resultado que arrojan estos gráficos es el siguiente:

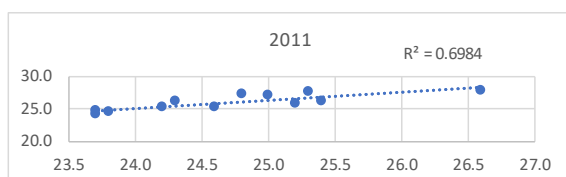
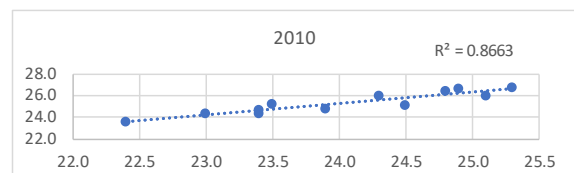
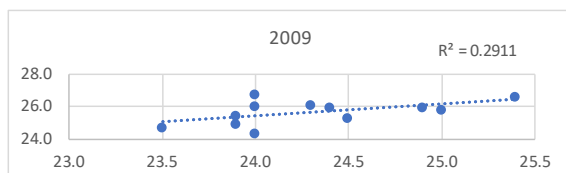
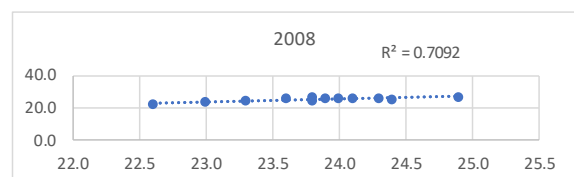
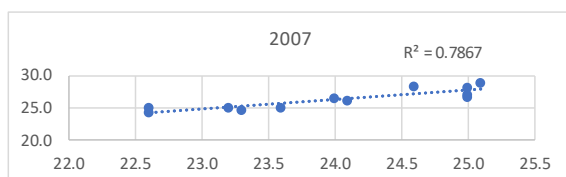
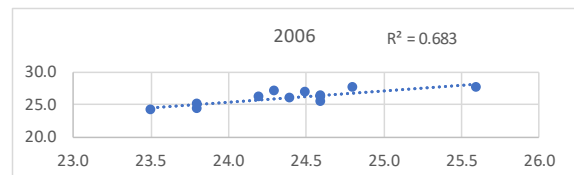
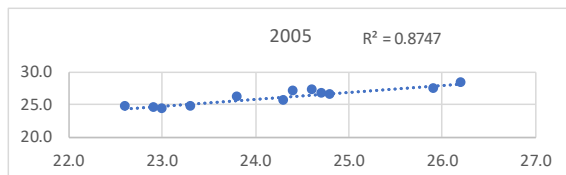
¹ Open-Meteo es una API meteorológica de código abierto y ofrece acceso gratuito para uso no comercial.
<https://open-meteo.com/>



2005	0.6192	2009	0.907
2006	0.9556	2010	0.4911
2007	0.6981	2011	0.6984
2008	0.9261		
PR.	0.7565		

Como se puede apreciar los coeficientes de determinación R^2 son bastante aceptables con excepción del año 2010 que desciende al valor de 0.4911, la razón por la cual se produce este descenso puede ser que sea que los datos registrados sean incorrectos o mal contabilizados o, que los datos de la API estén sobreestimados o subestimados, sea cual fuese el motivo de este descenso el valor promedio de los siete años considerados se ubica en 0.7565 lo que indica que existe una correlación positiva significativa.

En lo que respecta a la temperatura media del aire los resultados son los siguientes:



2005	0.8747	2009	0.2911
2006	0.683	2010	0.8663
2007	0.7867	2011	0.6984
2008	0.7092		
PR.	0.70134		

Para el caso de esta variable los coeficientes de determinación R^2 son menores a los obtenidos para los datos de precipitación sin embargo siguen siendo aceptables y aplicables ya que el valor promedio se ubica en 0.70134 lo que significa una correlación positiva significativa.

Metodología utilizada en el relleno de datos faltantes.

En el caso de los métodos para el relleno de datos ausentes se cuenta con las directrices de la Guía de Prácticas Climatológicas de la Organización Meteorológica Mundial (WMO, 1983), que entre otros, se proponen los siguientes métodos estadísticos para el relleno de datos faltantes: regresión simple, múltiple (Degaetano et al., 1995), razón q y razón-normal q (RN) (Paulhus y Kohler, 1952).

Para obtener los datos meteorológicos de la API producto del reanálisis de ERA5 y ERA5-Land (reanálisis atmosférico de quinta generación) producidos por el Centro Europeo de Previsiones Meteorológicas a Plazo Medio (ECMWF)

El reanálisis combina datos de modelos con observaciones de todo el mundo en un conjunto de datos globalmente completo y consistente utilizando las leyes de la física. El reanálisis produce datos que se remontan varias décadas atrás en el tiempo, proporcionando una descripción precisa del clima del pasado; en base a estos criterios se procede a generar año tras año los valores de cada una de las variables con el propósito de elaborar datos de la serie 1991-2020.

Una vez obtenidos los estos datos de la API se opta por utilizar el método de Regresión Ortogonal que consiste en que cuando se tiene series estocásticas que pueden correlacionarse entre sí y una de ellas tiene datos faltantes, se puede rellenar con datos de la serie que posee datos completos. A la serie completa (generada por la API) se le asigna el nombre de “Serie Índice” y a la incompleta (datos del INAMHI) “Serie de Análisis”; en términos generales este método minimiza la suma de las longitudes perpendiculares al cuadrado entre cada uno de los puntos de datos y la respectiva línea de regresión.

El método está basado en el análisis estadístico. En la correlación lineal la distancia de los puntos (observaciones) hasta la línea de tendencia se la toma paralela a los ejes X y Y , mientras que en la regresión ortogonal, dicha distancia es tomada perpendicular a la línea.

Se requiere, con la ayuda del método de los mínimos cuadrados, determinar cierto coeficiente λ , mediante la resolución de la ecuación cuadrática:

$$\lambda^2 - (\sigma_X^2 + \sigma_Y^2)\lambda + (\sigma_X^2\sigma_Y^2 - [\sigma_{XY}]^2) = 0$$

Se debe cumplir la condición:

$$\lambda_1 > \lambda_2 > 0$$

Como convencionalismo a la serie completa se la denomina “serie índice” y se la representa por X , y a la que le faltan datos, se la llama “serie incompleta”, a la que se le asigna la nominación Y .

La pendiente de la línea de regresión se la determina por:

$$m = \frac{\sigma_{XY}}{\lambda_1 - \sigma_Y^2}$$

Donde:

σ_X^2 y σ_Y^2 – Varianzas para las series X y Y

σ_{XY} – Covarianza

Finalmente, la fórmula para el cálculo de los datos faltantes es:

$$Y = a + bX$$

Donde:

a y b Coeficientes que se calculan con las fórmulas:

$$a = \bar{Y} - m\bar{X}$$

$$b = m$$

Donde:

$\bar{X}$ - Valor medio de las observaciones X

$\bar{Y}$ - Valor medio de las observaciones Y

En el caso cuando la línea de regresión interseca al eje de la X en el punto $X_0 > 0$, para todos $X < X_0$, los valores de las magnitudes Y serán negativos, por lo que para resolver esta contradicción, se recomienda utilizar para rellenar los datos, la expresión:

$$Y = \bar{Y} \left(\frac{X}{\bar{X}} \right)^{\left(\frac{m\bar{X}}{\bar{Y}} \right)}$$

Los cálculos estadísticos se realizarán en hojas excel como se muestra a continuación:

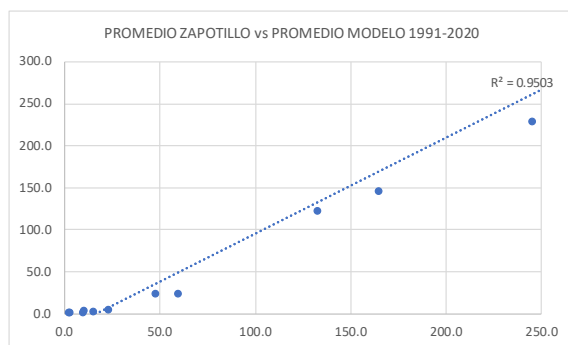
	ÍNDICE ANÁLISIS											
	MOD M0151 ENE	MOD M0151 FEB	MOD M0151 MAR	MOD M0151 ABR	MOD M0151 MAY	MOD M0151 JUN	MOD M0151 JUL	MOD M0151 AGO	MOD M0151 SEP	MOD M0151 OCT	MOD M0151 NOV	MOD M0151 DIC
1991	20.5	157.3	170.5	113.7	228.2	163.5	76.6	5.6	49.7	0.0	13.8	0.0
1992	177.2	36.3	179.0	95.9	587.9	392.8	604.5	172.0	270.2	155.5	6.9	0.0
1993	119.6	2.7	442.9	283.4	407.1	675.4	391.8	300.2	51.5	67.7	3.9	0.0
1994	204.0	269.4	362.4	403.7	234.9	288.7	258.1	288.1	51.9	27.1	17.7	0.0
1995	144.0	4.8	215.8	0.0	115.7	3.2	112.2	1.0	35.7	0.0	2.9	0.0
1996	107.2	5.8	139.3	39.7	188.4	38.6	57.2	3.4	31.3	0.0	5.1	0.0
1997	62.7	7.8	139.5	112.4	214.2	133.9	137.6	242.8	61.8	39.2	31.1	4.4
1998	385.9	619.4	547.0	568.5	662.3	685.6	424.0	883.7	73.2	73.5	17.3	0.0
1999	71.7	14.0	353.2	412.3	171.9	180.2	111.5	51.8	42.0	22.1	17.0	1.5
2000	46.7	96.4	163.9	205.8	228.7	373.2	142.7	157.3	35.7	23.7	18.6	7.1
2001	207.6	250.0	217.6	144.7	376.9	485.0	191.3	247.7	26.3	2.0	3.4	0.0
2002	42.2	55.2	293.5	84.3	385.2	945.7	337.6	280.9	25.2	12.6	6.2	0.0
2003	114.9	79.2	147.4	60.9	93.9	96.5	63.4	0.0	15.1	4.2	13.1	3.9
2004	65.7	141.1	198.1	50.1	122.4	28.6	108.5	25.5	19.4	6.9	1.4	0.0
2005	45.9	3.2	127.8	51.7	154.6	362.8	76.7	33.1	24.3	0.0	6.2	4.6
2006	90.5	105.9	279.2	568.2	217.6	394.0	115.2	132.7	17.2	0.0	12.9	2.0
2007	135.2	44.7	133.1	36.2	217.5	331.2	98.5	64.9	16.0	1.6	17.7	0.0
2008	164.4	185.5	340.8	463.9	360.3	683.4	187.6	150.7	38.9	5.0	21.5	16.9
2009	230.0	205.3	257.3	327.0	198.7	321.6	187.6	150.7	21.0	26.6	7.7	0.0
2010	144.9	55.8	409.1	298.7	169.6	537.4	151.9	139.1	35.2	22.0	13.1	14.0
2011	104.6	10.7	68.1	98.1	48.9	7.5	175.4	211.5	16.7	13.1	12.4	15.3
2012	218.1	302.2	532.5	680.8	263.7	348.7	203.4	201.3	61.9	33.3	2.3	1.9
2013	81.3	4.7	153.9	64.1	271.7	365.3	50.0	9.4	45.0	36.8	0.4	0.0
2014	180.8	27.9	83.3	48.3	164.4	118.6	43.2	6.8	112.9	65.1	9.4	2.4
2015	116.8	19.9	134.9	43.7	361.0	450.1	140.3	83.3	101.0	28.6	16.0	10.2
2016	199.9	382.1	281.7	272.3	323.1	456.3	161.1	134.2	15.9	4.7	4.2	0.2
2017	157.1	165.0	346.6	378.0	801.4	786.2	233.9	249.7	68.7	37.6	8.0	1.8
2018	80.4	7.0	90.5	86.8	77.4	55.6	77.2	45.1	23.1	9.2	4.0	0.4
2019	175.6	203.3	426.8	508.6	167.0	147.4	61.4	23.9	29.1	12.9	7.0	1.4
2020	87.5	84.9	132.8	29.8	110.9	30.6	94.0	27.8	23.0	9.1	2.9	0.4

Las celdas sombreadas significan valores obtenidos durante el proceso de cálculo.

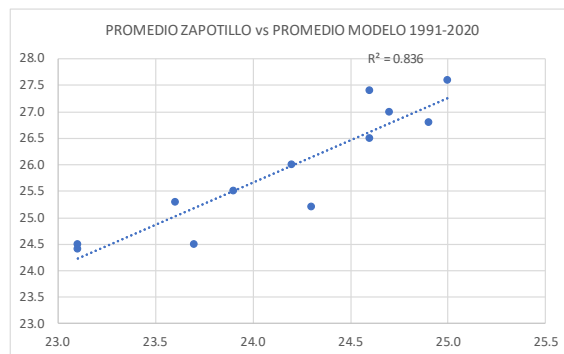
	ÍNDICE ANÁLISIS											
	MOD M151 ENE	MOD M151 FEB	MOD M151 MAR	MOD M151 ABR	MOD M151 MAY	MOD M151 JUN	MOD M151 JUL	MOD M151 AGO	MOD M151 SEP	MOD M151 OCT	MOD M151 NOV	MOD M151 DIC
ACIONES:												
SX=	75.90	124.26	155.10	140.73	55.71	7.88	3.51	2.59	22.71	17.13	22.41	88.51
SY=	150.41	178.83	259.64	188.10	37.30	5.48	0.92	1.05	5.38	2.62	6.89	73.69
IANZAS:												
Sx2=	5760.51	#####	#####	#####	3103.25	62.02	12.29	6.71	515.71	293.33	502.02	7834.47
Sy2=	#####	#####	#####	#####	1391.22	30.07	0.85	1.10	28.94	6.85	47.52	5430.12
IRIANZA:												
Sxy=	8897.95	#####	#####	#####	1739.60	16.04	0.70	0.83	109.82	34.56	103.99	5957.26
l/Xmed=	0.79489	0.81477	1.30542	0.86405	0.52068	0.28162	0.09556	0.12633	0.112003	0.10699	0.14785	0.40967
IENTES:												
a=	-204.942	-186.520	-200.180	-121.317	-5.179	-1.512	0.118	-0.047	-1.224	-0.207	-1.834	-24.442
b=	2.325	1.629	2.081	1.586	0.622	0.415	0.061	0.145	0.215	0.119	0.218	0.818
A	1	1	1	1	1	1	1	1	1	1	1	1
B	#####	#####	#####	#####	-4494.470	-92.088	-13.133	-7.813	-544.653	-300.175	-549.541	#####
C	#####	#####	#####	#####	#####	1607.55	9.93	6.70	2865.58	814.73	13041.26	#####
da l: Ll=	#####	#####	#####	#####	4186.038	68.683	12.328	6.832	539.339	297.436	524.686	#####
nte: m=	2.325	1.629	2.081	1.586	0.622	0.415	0.061	0.145	0.215	0.119	0.218	0.818

El mismo proceso se realizó para el tratamiento estadístico de la serie de temperatura media en base a los datos de la serie de la estación M0151 Zapotillo y los generados a través de la API meteorológica.

Una vez realizado el relleno de los datos faltantes en la serie M0151 tanto para los datos de precipitación como de temperatura media se obtuvo los promedios mensuales para el período en consideración; los valores promedios de serie fueron comparados con los de la serie de datos de la API para lo cual se procedió a la elaboración de gráficos de dispersión para obtener los coeficientes de determinación R² cuyos resultados se aprecian a continuación:



PRECIPITACIÓN



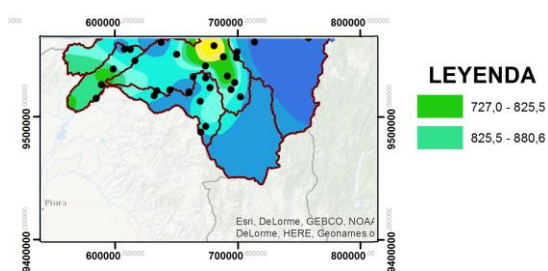
TEMPERATURA MEDIA

Como puede apreciarse, en base a los resultados obtenidos el valor de 0.9503 para la precipitación indica una correlación positiva fuerte mientras que el valor de 0.836 para la temperatura media significa una correlación positiva significativa.

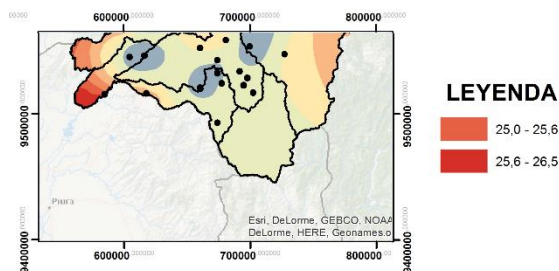


En tal sentido y en base a los resultados obtenidos, pese a que el porcentaje de datos faltantes de la estación de Zapotillo es alto, sin embargo los datos generados mediante el uso de la API para ser aplicados en el método estadístico aquí descrito dan como resultado de que se los puede considerar como fiables, permitiendo de esta manera realizar el relleno de los datos faltantes.

Complementariamente se realizó una comprobación del valor promedio multianual de precipitación de la serie 1991-2020 (894.0 m.m.) con el valor del mapa de isoyetas medias anuales de la serie histórica 1981-2010 mismo que se ubica en la cota de entre los 825.5 y 880.6 m.m., la diferencia puede explicarse en que los períodos considerados son diferentes pero muestran valores muy cercanos entre sí; en cuanto a la temperatura media el valor promedio de la serie 1991-2020 es de 25.9°C mientras que el indicado en el mapa de isotermas medias anuales del período 1981-2010 se ubica entre los 25.6 y 26.5°C que de la misma manera puede explicarse por la diferencia de los períodos considerados.



MAPA DE ISOYETAS 1981-2010



MAPA DE ISOTERMAS 1981-2010

Finalmente, en el Reglamento Técnico (OMM, 2016b) se facilitan las siguientes definiciones:

- **Medias de períodos:** Medias de los datos climatológicos calculadas para cualquier período de por lo menos 10 años que comience el 1 de enero de un año que acabe en 1.
- **Normales:** Medias periódicas calculadas para un período uniforme y relativamente largo que comprenda por lo menos tres períodos consecutivos de 10 años.
- **Normales climatológicas reglamentarias:** Medias de los datos climatológicos calculadas para los siguientes períodos consecutivos de 30 años: 1 de enero de 1981 a 31 de diciembre de 2010, 1 de enero de 1991 a 31 de diciembre de 2020, y así sucesivamente.

Partiendo de estas definiciones los valores obtenidos de la estación de Zapotillo corresponderían a una Normal Climatológica Reglamentaria.

Los datos proporcionados por el INAMHI así como los generados por la API y el relleno realizado luego de la implementación del método de Regresión Ortogonal se adjunta en un archivo Excel para su respectivo uso.

Atentamente,

MSc. Humberto Enríquez D.

Meteorólogo